Computational approach to the emergence and evolution of language - evolutionary naming game model

Adam Lipowski[1] and Dorota Lipowska[2]

[1] Faculty of Physics, Adam Mickiewicz University, Poznań 61-614, Poland
   lipowski@amu.edu.pl
[2] Institute of Linguistics, Adam Mickiewicz University, Poznań 60-371, Poland
   lipowska@amu.edu.pl

Summary. Computational modelling with multi-agent systems is becoming an important technique of studying language evolution. We present a brief introduction into this rapidly developing field, as well as our own contributions that include an analysis of the evolutionary naming-game model. In this model communicating agents, that try to establish a common vocabulary, are equipped with an evolutionarily selected learning ability. Such a coupling of biological and linguistic ingredients results in an abrupt transition: upon a small change of the model control parameter a poorly communicating group of linguistically unskilled agents transforms into almost perfectly communicating group with large learning abilities. Genetic imprinting of the learning abilities proceeds via Baldwin effect: initially unskilled communicating agents learn a language and that creates a niche in which there is an evolutionary pressure for the increase of learning ability. Under the assumption that communication intensity increases continuously with finite speed, the transition is split into several transition-like changes. It shows that the speed of cultural changes, that sets an additional characteristic timescale, might be yet another factor affecting the evolution of language. In our opinion, this model shows that linguistic and biological processes have a strong influence on each other and this effect certainly has contributed to an explosive development of our species.

Key words: computer modelling, agent systems, naming game, Baldwin effect, bio-linguistics

1 Introduction

1.1 Evolutionary forces behind language development

The ability to use language distinguishes humans from all other species. Certain species also developed some communication modes but of much smaller capabilities as well as complexity. Since several decades various schools are

trying to explain the emergence and development of language. Nativists argue that langauge capacity is a collection of domain-specific cognitive skills that are somehow encoded in our genome. However, the idea of the existence of such a Language Acquisition Device or "language organ" (the term coined by their most prominent representative Noam Chomsky [1]), was challenged by empiricists, who argue that linguistic performance of humans can be explained using domain-general learning techniques. The recent critique along this line was made by Sampson [2], who questions even the most appealing argument of nativists, that refer to the poverty of stimulus and apparently fast learning of grammar by children. An important issue of possible adaptative merits of language does not seem to be settled either. Non-adaptationists, again with Chomsky as the most famous representative [3], consider language as a side effect of other skills and thus claim that its evolution, at least at the beginning, was not related with any fitness advantage. A chief argument against the non-adaptationist stand is the observation that there is a number of costly adaptations that seem to support human linguistic abilities such as a large brain, a longer infancy period or descended larynx. Recently, in their influential paper Pinker and Bloom argued that, similarly to other complex adaptations, language evolution can only be explained by means of natural selection mechanisms [4]. Their paper triggered a number of works where language was examined from the perspective of evolutionary biology or game theory [5, 6]. In particular, Nowak et al. used some optimization arguments, that might explain the origin of some linguistic universals [7]. They suggest that words appeared in order to increase the expressive capacity and sentences (made of words) limit memory requirements. Confrontation of nativists with empiricists and adaptationists with non-adaptationists so far does not seem to lead to consensus but certainly deepened our understanding of these problems [8].

Recently, a lot of works on the language emergence seem to have an evolutionary flavour. Such an approach puts some constraints on possible theories of the language origin. In particular, it rules out non-adaptationist theories, where language is a mere by-product of having a large and complex brain [9]. The emergence of language has been also listed as one of the major transitions in the evolution of life on Earth [10]. An interesting question is whether this transition was variation or selection limited [11]. In variation limited transitions the required configuration of genes is highly unlikely and it takes a considerable amount of time for the nature to invent it. For selection limited transitions the required configuration is easy to invent but there is no (or only very weak) evolutionary pressure that would favour it. Relatively large cognitive capacities of primates and their genetic proximity with humans suggests that some other species could have been also capable to develop language-like communication. Since they did not, it was perhaps due to a weak selective pressure. Such indirect arguments suggest that the emergence of language was selection limited [11].

Some interesting results can be obtained by applying game-theory reasoning to one of the most basic problems of emerging linguistic communication, namely why do we talk and why do we exchange valuable and trustful information. Since speaking is costly (it takes time, energy and sometimes might expose a speaker to predators), and listening is not, such a situation seems to favour selfish individuals that would only listen but would not speak. Moreover, in the case of the conflict of interests the emerging communication system would be prone to misinformation or lying. The resolution of these dilemmas usually refers to the kin selection [12] or reciprocal altruism [13]. In other words, speakers remain honest because they are helping their relatives or they expect that others will do the same for them in the future. As an alternative explanation Dessalles [14] suggests that honest information is given freely because it is profitable - it is a way of competing for status within a group. Some related results on computational modelling of the honest cost-free communication are reported by Noble [15].

A necessary ingredient of language communication is learning. It is thus legitimate to ask whether darwinian selection might be responsible for the genetic hard-wiring of a Language Acquisition Device. Indeed, this (to some extent hypothetical) organ is most likely responsible for some of the arbitrary (as opposed to the functional) linguistic structures. But for such an organ to be of any value, an individual has to acquire the language first. The inheritance of characteristics acquired during an individual lifetime is usually associated with discredited lamarckian mechanism and thus considered to be suspicious. However, the relation between evolution and learning is more delicate and the attempts to clarify the mutual interactions of these two adaptive mechanisms have a long history. According to a purely darwinian explanation, known as a Baldwin effect [16, 17, 18], there might appear a selective pressure in a population for the evolution of the instinctive behaviour that would replace the beneficial, but costly, learned behaviour [19]. Baldwin effect presumably played an important role in the emergence and evolution of language but certain aspects of these processes still remain unresolved [20]. For example, one of the assumptions that is needed for the Baldwin effect to be effective is a relatively stable environment since otherwise rather slow evolutionary processes will not catch up with the fast changing environment. Since the language formation processes are rather fast (in comparison to the evolutionary time scale), Christiansen and Chater questioned the role of adaptive evolutionary processes in the formation of arbitrary structures like Language Acquisition Device [21]. Actually, they suggest a much different scenario, where it is a language that adapted to human brain structures rather than vice versa.
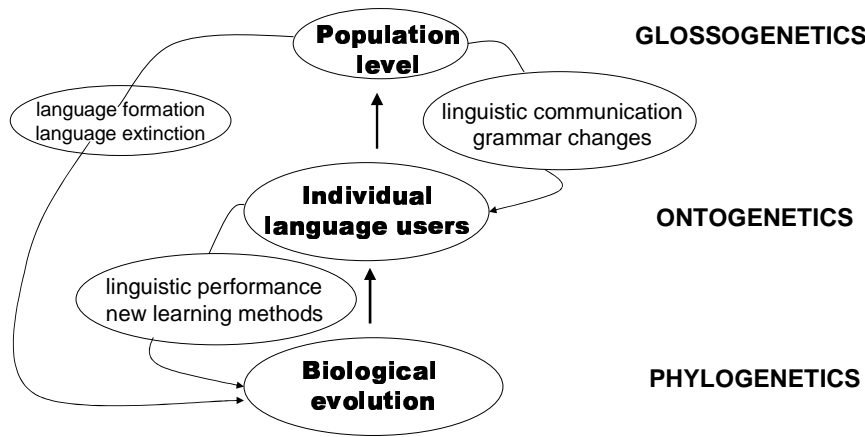
1.2 Language as a complex adaptive system

From the above description it is clear that studying of the emergence and evolution of language is a complex and multidisciplinary task and requires cooperation of not only linguists, neuroscientists, and anthropologists, but

also experts in artificial intelligence, computer sciences or evolutionary biology [22]. One can distinguish two levels at which language can be studied and described [23] (Fig. 1). At the individual level the description centers on the individual language users: their linguistic performance, language acquisition, speech errors, speech pathologies or brain functioning in relation with language processing. At the individual level the language of each individual is slightly different. Nevertheless, within certain population these individuals can efficiently communicate and that establishes the population level. At this level the language is considered as an abstract system that exists in a sense separately from the individuals users. There are numerous interactions between these two levels. Indeed, the linguistic behaviour of individuals depends on the language (at the population level) specific to the population they are part of. And, as a feedback, the language used in a given population is a collective behaviour and emerges from linguistic behaviour of individuals composing this population. Various processes shaping such a complex system are operating at different time scales. The fastest dynamics is operating at the individual level (ontogenetic timescale [24]) that includes, for example, language acquisition processes. Much slower processes, such as migrations of language populations, dialects formation or language extinctions, are operating at the so-called glossogenetic timescale. The slowest processes govern the biological evolution of language users and that defines the phylogenetic timescale. Processes operating at these different timescales are not independent (Fig. 1). Biological evolution might change linguistic performance of individuals and that might affect the glossogenetic processes. For example, a mutation that changes the vocal ability of a certain individual, if spread in his/her population, might lead to a dialect formation or a language extinction. Such population-level processes might change the selective pressure that individual language users are exposed to and that might affect phylogenetic processes, closing thus the interaction loop.

Various levels of descriptions and processes operating at several timescales suggest that complex models must be used to describe adequately the language evolution. Correspondingly, the analysis of such models and predicting their behaviour also seem to be difficult. It is known that some phenomena containing feedback interactions might be described in terms of nonlinear differential equations, such as, for example, Lotka-Volterra equations describing interacting populations. The behaviour of such nonlinear equations is often difficult to predict, since abrupt changes even of the qualitative nature of solutions might take place. Language evolution is, however, much more complex than ecological problems of interacting populations and its description in terms of differential equations would be much more complicated if at all feasible.

Recently, it seems that the most promising and frequently used approach to examine such systems is computational modelling of multi-agent systems. Using this method one examines a language that emerges in a bottom-up fashion as a result of interactions within a group of agents equipped with some

Fig. 1. Language as a complex adaptive system.Many different processes governing the language evolution are entangled at various levels. Relatively fast individual level (ontogenetics), comprising e.g., langauge acquisition processes, is determined mainly by interactions between individual language users. Much slower are populational-level processes (glossogenetics) such as language formations, extinctions, grammar changes or migrations. To obtain a complete description one has to consider also biological evolution (phylogenetics) and these are the slowest processes of the language evolution. Various processes at individual and population level affect the fitness landscape and that influences the biological evolution level. Similarly, individual language user level is affected by populational level processes.

linguistic functions. Then one considers language as a complex adaptive system that evolves and complexifies according to biologically inspired principles such as selection and self-organization [25]. Thus, the emerging language is not static but evolves in a way that hopefully is similar to human language evolution. Of course, using such an approach one cannot explain all intricacies of human languages. A more modest goal would be to understand some rather basic features that are common to all languages such as meaning-form mappings, origin of linguistic coherence (among agents without central control and global view), or coevolutionary origin of grammar and meaning.

Within such a multi-agent approach, two groups of models can be distinguished. In the first one, originating from the so-called iterated learning model, one is mainly concerned with the transmission of language between successive

generations of agents [26, 27]. Agents that are classified as teachers produce some expressions that are passed to learners that try to infer their meaning using statistical learning techniques such as neural networks. After a certain number of iterations teachers are replaced by learners and a new population of learners is introduced. The important issue that the iterated learning model has successfully addressed is the transition from holistic (complex meaning expressed by a single form) to compositional language (composite meaning is expressed with composite form). However, since such a procedure is computationally relatively demanding and the number of communicating agents is thus typically very small, the problem of the emergence of linguistic coherence must be neglected in this approach. To tackle this problem Steels introduced a naming game model [28]. In this approach one examines a population of agents trying to establish a common vocabulary for a certain number of objects present in their environment. The change of generations is not required in the naming game model since the emergence of a common vocabulary is a consequence of the communication processes between agents, and agents are not divided into teachers and learners but take these roles in turn.

It seems that the iterated learning model and the naming-game model are at two extremes: the first one emphasizes the generational turnover while the latter concentrates on the single-generation (cultural) interactions. Since in the language evolution both aspects are present, it is desirable to examine models that combine evolutionary and cultural processes. Recently we have introduced such a model [29] and one of the objectives of the present paper is to provide further analysis of its behaviour based on more extensive simulations. Our model captures all three basic aspects of language: learning, culture, and evolution. In our model agents try to establish a common vocabulary like in the naming game model, but in addition they can breed, mutate, and die. Moreover, they are equipped with an evolutionary trait: learning ability. When communication between agents is sufficiently frequent, cultural processes create a niche in which a larger learning ability becomes advantageous. It causes an increase of learning ability, but its large value in turn makes the cultural processes more efficient. As a result the model was shown to undergo an abrupt bio-linguistic transition where both linguistic performance and ability of agents change very rapidly [29]. One of the main results reported in this paper is that under the plausible assumption, that the intensity of communication increases continuously in time, this bio-linguistic transition is replaced with a series of fast, transition-like changes. In our opinion, the proposed model shows that linguistic and biological processes have a strong influence on each other, which has certainly contributed to an explosive development of our species. That learning in our model modifies the fitness landscape of a given agent and facilitates the genetic accommodation of learning ability is actually a manifestation of the much debated Baldwin effect.

2 Model

In our model we consider a set of agents located at sites of the square lattice of the linear size L. Agents are trying to establish a common vocabulary on a single object present in their environment. An assumption that agents communicate only on a single object does not seem to restrict the generality of our considerations and has already been used in some other studies of naming game [30, 31] or language-change [32] models. A randomly selected agent takes the role of a speaker that communicates a word chosen from its inventory to a hearer that is randomly selected among nearest neighbours of the speaker. The hearer tries to recognize the communicated word, namely it checks whether it has the word in its inventory. A positive or negative result translates into communicative success or failure, respectively. In some versions of the naming game model [30, 31] a success means that both agents retain in their inventories only the chosen word, while in the case of failure the hearer adds the communicated word to its inventory.

To implement the learning ability we have modified this rule and assigned weights w_i ($w_i > 0$) to each i-th word in the inventory. The speaker selects then the i-th word with the probability $w_i / \sum_j w_j$ where summation is over all words in its inventory (if its inventory is empty, it creates a word randomly). If the hearer has the word in its inventory, it is recognized. In addition, each agent k is characterized by its learning ability l_k ($0 < l_k < 1$), that is used to modify weights. Namely, in the case of success both speaker and hearer increase the weights of the communicated word by their learning abilities, respectively. In the case of failure the speaker subtracts its learning ability from the weight of the communicated word. If after such a subtraction a weight becomes negative, the corresponding word is removed from the repository. The hearer in the case of failure, i.e., when it does not have the word in its inventory, adds the communicated word to its inventory with a unit weight.

Apart from communication, agents in our model evolve according to the population dynamics: they can breed, mutate, and eventually die. To specify intensity of these processes we have introduced the communication probability p. With the probability p the chosen agent becomes a speaker and with the probability $1 - p$ a population update is attempted. During such a move the agent dies with the probability $1 - p_{\mathrm{surv}}$, where $p_{\mathrm{surv}} = \exp(-at)[1 - \exp(-b \sum_j w_j / \langle w \rangle)]$, and $a \sim 0.05$ and $b = 5$ are certain parameters whose role is to ensure a certain speed of population turnover. Moreover, t is the age of an agent and $\langle w \rangle$ is the average (over agents) sum of weights. Such a formula takes into account both its linguistic performance (the bigger $\sum_j w_j$ the larger p_{surv}) and its age. If the agent survives (it happens with the probability p_{surv}), it breeds, provided that there is an empty site among its neighbouring sites. The offspring typically inherits parent's learning ability and the word from its inventory that has the highest weight. In the offspring's inventory the weight assigned initially to this word equals one. With the small probability p_{mut} a mutation takes place and the learning ability of an offspring is selected

randomly anew. With the same probability an independent check is made whether to mutate the inherited word. A diagram illustrating the dynamics of our model is given in the Appendix [33]. Let us also notice that the behaviour of our model, that is described below, is to some extent robust with respect to some modifications of its rules. For example, qualitatively the same behaviour is observed for modified parameters a and b, different form of the survival probability p_{surv} (provided it is a decreasing function of t and an increasing function of $\sum_j w_j$), or different breeding and/or mutation rules.

3 Results

To examine the properties of the model we used numerical simulations. Most of the results are obtained for $L = 60$ and $p_{\mathrm{mut}} = 0.001$ but simulations for $L = 80$ and $p_{\mathrm{mut}} = 0.01$ lead to a similar behaviour. Simulations typically start from all sites occupied by agents that have a single word in their inventories, that is chosen randomly for each agent and assigned a unit weight. The learning ability of each agent is also chosen randomly.

3.1 Bio-linguistic transition

An important parameter of the model is the communication probability p that specifies the intensity of communication attempts in comparison with populational changes. In general, for small p the model remains in the phase of linguistic disorder with only small clusters of agents using the same language. We define the language of an agent as the largest-weight word in its inventory. Such a definition means that agents using the same language usually (but not always) use a recognizable word and it ensures a relatively large rate of communication successes for such agents. A typical distribution of languages in this disordered small-p phase is shown in the left panel of Fig. 2, where agents using the same language are drawn with the same shade of grey. Upon increasing the communication probability p the clusters of agents only slightly increase, but after reaching a certain threshold an abrupt transition takes place and the model enters the phase of linguistic coherence with almost all agents belonging to the same cluster (Fig. 2, right panel). To examine the nature of this transition we have measured the communication success rate s defined as an average over agents and simulation time of the fraction of successes with respect to all communication attempts (Fig. 3). Moreover, we have measured the average learning ability l (Fig. 4). One can notice that upon increasing p the abrupt transition takes place around $p = 0.23$, where both the communication success rate s and the learning ability l jump. Moreover, upon decreasing p this transition takes place at a much lower value, namely around $p = 0.15$. Such a hysteretic behaviour indicates that the transition in our model is discontinuous. We also examined the behaviour of the model with the learning ability kept fixed during entire simulations. In this case there is

)

p=0.15

p=0.3

language   learning ability
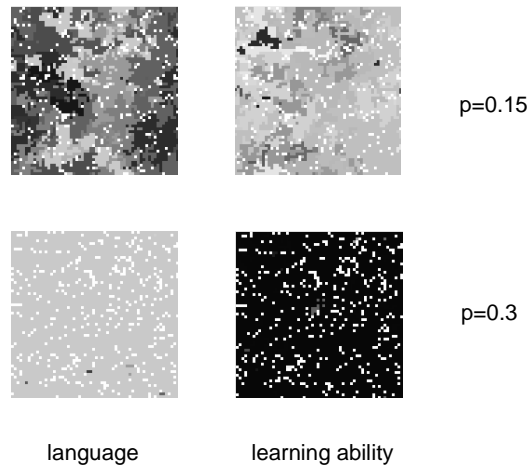
Fig. 2. Exemplary configurations of the evolutinary naming game model with $L = 60$ and $p_{\mathrm{mut}} = 0.001$. In the small-p phase (upper panel) communications are infrequent and agents using the same language (left)or having the same learning abilities (right) form only small clusters. In this phase the communication success rate s and the learning ability l are small (see also Figs. 3-4). The larger the learning ability of an agent the darker are pixels representing it (white: l=0; black: l=1). In the large-p phase (lower panel) frequent communications result in the emergence of the common language. Moreover, almost all agents use the same language and have the same, and large, learning ability.

also a phase transition between disordered and linguistically coherent phases but this time the transition is much smoother and there is no indication of the hysteretic behaviour (Fig. 3). To get further insight into the behaviour of our model, we have measured the fraction f_m of agents using the language with the largest number of users. Simulations show that for the learning ability kept fixed f_m also decreases in a much smoother way (Fig. 5). Moreover, its variance has a pronounced peak at the transition point that this time takes place around $p = 0.07$ (Fig. 6). Such large fluctuations of f_m (and a similar behaviour shows the variance of s) in the vicinity of the transition point and an absence of the jump suggest that this might be a continuous transition. In the last section we will return to this point.

A noticeable difference between small-p and large-p phases appears in the learning-ability dependence of lifetime of agents (Fig. 7). One can see that in the large-p phase, where neighbouring agents are likely to use the same language, having a large learning ability increases the agent's lifetime (faster learning enables faster accommodation to the predictable linguistic environment). On the other hand, in the small-p phase (i.e., in the random linguistic environment) the lifetime is almost independent on the learning ability. Before
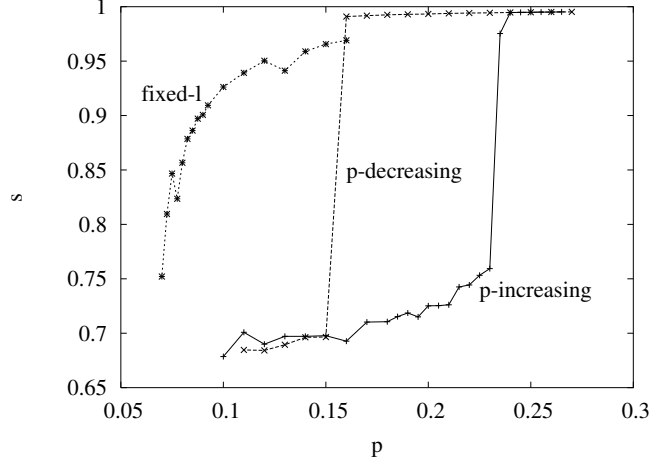
Fig. 3. The success rate s as a function of the communication probability p. Calculations were made for system size $L = 60$ and mutation probability $p_{\mathrm{mut}} = 0.001$. Simulation time for each value of p was typically equal to 10^5 steps with $3 \cdot 10^4$ steps discarded for relaxation. A step is defined as a single, on average, update of each site. For simulations with decreasing p we first relaxed the system until a mono-language state was reached (with s and l close to unity). In the left part of the graph the data are from simulations with fixed $l (= 0.5)$.

presenting computational results concerning the dynamics of our model, let us notice that sudden transitions in linguistic models were also reported in some other models [7, 38].

3.2 Dynamic behaviour

Because each agent is characterized by its learning ability, homogeneous states, namely states where a majority of agents are using the same language, but of different learning abilities are not equivalent. As a result, evolution of the model in an intricate way depends on the initial configuration and the parameters p and p_{mut}. This is particularly transparent in the range $0.15 < p < 0.25$, where the model exhibits hysteretic behaviour (for $p_{\mathrm{mut}} = 0.001$). An example that shows the dependence on the initial configuration is shown in Fig. 8. In this case inside $L = 60$ lattice we have created a square seed of 100 agents having the same learning ability 0.98 and the same word in their inventories. This seed is surrounded by $60 \cdot 60 - 100 = 3500$ agents of smaller learning ability 0.5. As can be seen in Fig. 8, the evolution depends on whether the
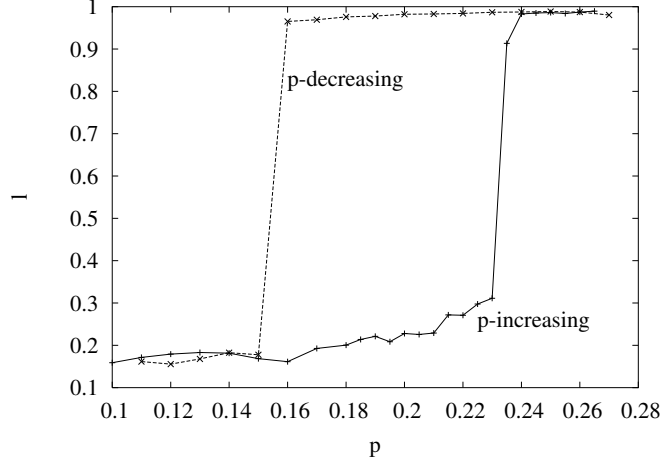
Fig. 4. The average learning ability l as a function of the communication probability p. Details of the simulations are the same as in Fig. 3

surrounding agents are using the same language as those in the seed (homogenous) or whether initially their repositories contain random words (random). In the first case the system ends up in the homogenous state where a majority of agents are using the same language and have the same learning ability (0.98). In the second case the model evolves toward the multi-language state with much smaller learning abilities. In such a setup the size of the seed or the learning ability of the surrounding agents are also important parameters that might affect the course of the evolution of the model. For example, we observed that in the homogenous case but for surrounding agents having the learning ability 0.3 the model evolved toward the multi-language state.

3.3 Baldwin effect in time varying environment

The fact that the success rate s and the learning ability l have a jump at the same value of p (Figs. 3-4) shows that communicative and biological ingredients in our model strongly influence each other and that leads to the single and abrupt transition. In our model successful communication requires learning. A new-born agent communicating with some mature agents who already worked out a certain (common in this group) language will increase the weight of a corresponding word. As a result, in its future communications the agent will use mainly this word. In what way such a learning might get coupled with evolutionary traits? The explanation of this phenomenon is known as a
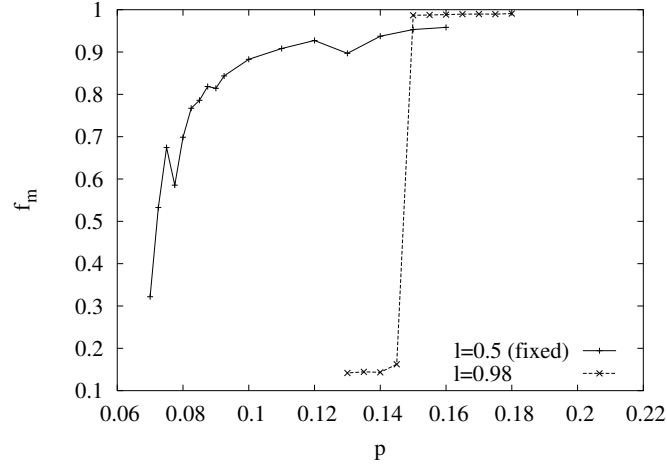
Fig. 5. The fraction f_m of agents using the language with the largest number of users as a function of p. For simulations with the learning ability not kept fixed we started from the configuration with all agents having the same word in their repositories and the learning ability set to 0.98. Such a choice of an initial state leads to only minor differences with simulations in Figs. 3-4
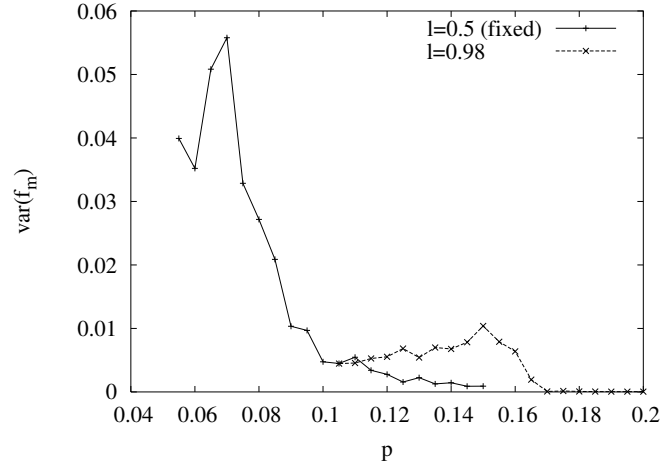
Fig. 6. The variance of f_m. Details of simulations are the same as in Fig. 5.
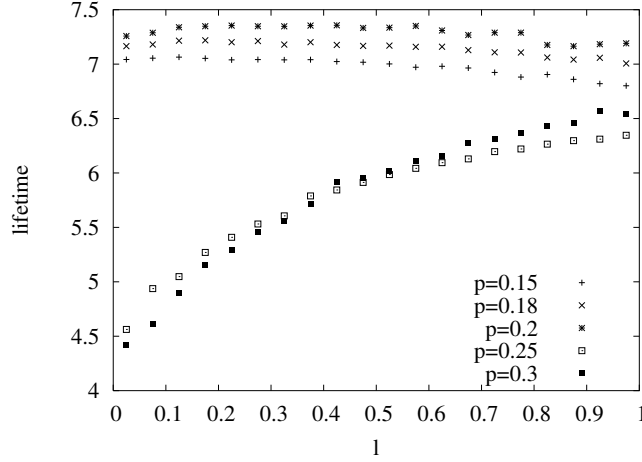
Fig. 7. Lifetime of agents as a function of learning ability l for several values of communication probability p. One can notice that in the predictable environment (large-p phase) having large learning ability is of advantage. In the random environment (low-p phase), lifetime of an agent is almost independent on its learning ability.

Baldwin effect. Although at first sight it looks like a discredited Lamarckian phenomenon, the Baldwin effect is actually purely Darwinian [34, 20]. There are usually some benefits related with the task a given species has to learn and there is a cost of learning this task. One can argue that in such case there is some kind of an evolutionary pressure that favours individuals for which the benefit is larger or the cost is smaller. Then, the evolution will lead to the formation of species where the learned behaviour becomes an innate ability. It should be emphasized that the acquired characteristics are not inherited. What is inherited is the ability to acquire the characteristics (the ability to learn) [19]. In the context of the language evolution the importance of the Baldwin effect was suggested by Pinker and Bloom [4]. Perhaps this effect is also at least partially responsible for the formation of the Language Acquisition Device. However, many details concerning the role of the Baldwin effect in the evolution of language remain unclear [35].

We already argued [29], that in our model the Baldwin effect is also at work. Let us consider a population of agents with the communication probability p below the threshold value ($p = p_c \approx 0.23$). In such a case the learning ability remains at a rather low level (since clusters of agents using the same language are small, it does not pay off to be good at learning the language
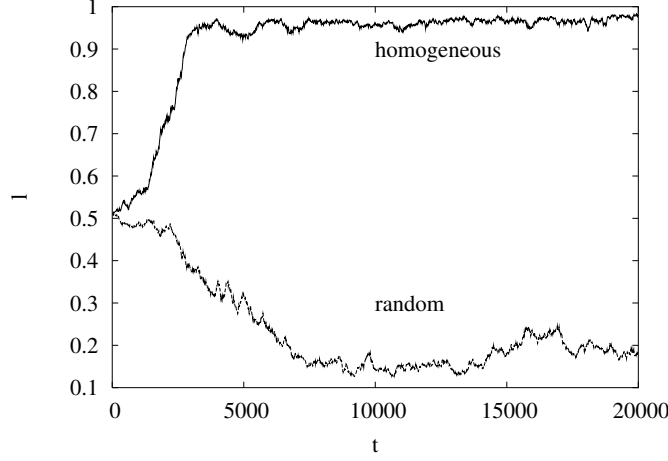
Fig. 8. Time evolution of the learning ability l for the $L = 60$ model with the seed of 100 agents with $l = 0.98$ surrounded by 3500 agents with $l = 0.5$. The course of the evolution depends on the initial state (inventories) of surrounding agents (see main text for the detailed description).

of your neighbours). Now, let us increase the value of p above the threshold value. More frequent communication changes the behaviour dramatically. Apparently, clusters of agents using the same language are now sufficiently large and it pays off to have a large learning ability because that increases the success rate and thus the survival probability p_{surv}. Let us notice that p_{surv} of an agent depends on its linguistic performance ($\sum_j w_j$) rather than its learning ability. Thus clusters of agents of good linguistic performance (learned behaviour) can be considered as niches that direct the evolution by favouring agents with large learning abilities, which is precisely the Baldwin effect. It should be noticed that linguistic interactions between agents (whose rate is set by the probability p) are typically much faster than evolutionary changes (set by p_{mut}) and such an effect was observed in simulations [29].

As a result of a positive feedback (large learning ability enhances communication that enlarges clusters that favours even more the increased learning ability) a discontinuous transition takes place both with respect to the success rate and learning ability (Figs. 3-4). An interesting question is whether such a behaviour is of any relevance in the context of human evolution. It is obvious that development of language, which probably took place somewhere around 10^5 years ago, was accompanied by important anatomical changes such as fixation of the so-called speech gene (FOXP2), descended larynx or

enlargement of brain [36]. Linguistic and other cultural interactions that were already emerging in early hominid populations were certainly shaping the fitness landscape and that could direct the evolution of our ancestors via the Baldwin effect.

Since it is plausible that communication attempts in the human history were gradually becoming more frequent (and important), it is natural to simulate our model with the communication probability p increasing continuously in time. Because with respect to linguistic abilities human population is rather homogeneous, it would be desirable that in our model dynamics would arrive at an l-homogeneous state, i.e., a state where majority of agents have the same, and rather large, linguistic abilities (we expect that in our model the linguistic ability l, as a heritable and unchangeable during a phenotypic development feature, approximately corresponds to the Language Acquisition Device). In the initial population of agents the learning ability should be rather low. Although many languages are now at the verge of extinction and one cannot exclude that in the future humans will use only one language, at least at present many languages exist. Thus we expect that in the final or at least in transient but long-lived state there will be many (or several, taking into account limitations of the simulated systems) languages. Results of the simulations with such a setup are seen in Fig. 9. Initially, we set the learning abilities as random numbers uniformly distributed from the interval (0,0.1). One can notice that around $t = 5 \cdot 10^4$ a learning ability close to 0.1 dominates in the system ($f_l \approx 1$). However, upon an increase of time (and p), such a low learning ability is not sufficient and around $t = 12 \cdot 10^4$ the system becomes dominated by agents with learning ability close to 0.3. But still f_m remains close to 0, which means that even the most abundant language in the system is used only by a few agents. Around $t = 15 \cdot 10^4$ the next transition takes place and a large learning ability dominates in the system. Around that time f_m starts to increase and that means that some languages start to grow and some get extinct. Since almost all agents have the same language ability, all languages are dynamically equivalent, and this stage resembles domain coarsening in, for example, the Potts model (in Conclusions we argue, however, that there might be some differences in the behaviour of our model and the Potts model). Eventually, the system reaches the state where almost all agents use the same language $f_m, s \approx 1$, however, the time needed to reach such a state might be quite long.

In Fig. 9 the behaviour of the model in the interval $15 \cdot 10^4 < t < 2 \cdot 10^5$ resembles the current stage of evolution of human language: a single learning ability dominating the entire population and several (not too many and not too few) languages in use. Before arriving at such a state some plateaus can be distinguished separated with relatively rapid transitions. Such a behaviour differs from the single-step scenario seen in the simulations where p increases in finite steps but is kept constant during measurements (Figs. 3-4). Presumably, a multi-step behaviour is a consequence of the finite-speed increase of p. Let us notice that basic factors that determine the evolution of language set some
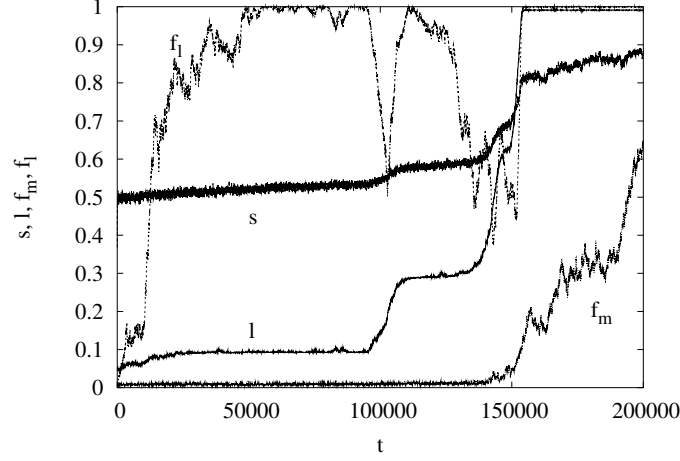
Fig. 9. Time evolution of the model characteristics upon the linear in time increase of the communication probability p from 0.1 to 0.5 ($L = 60$). We measured the success rate s, the learning ability l, the fraction of agents using the language with the largest number of users f_m, and the fraction of agents having the most abundant learning ability f_l. One can see that around $t = 15 \cdot 10^4$ both f_l and l becomes close to unity which means that almost every agent have the same and large learning ability. Further evolution gradually eliminates less abundant languages and leads to the state where almost all agents use the same language ($f_m, s \approx 1$).

characteristic timescales of the corresponding processes. Namely, individual learning - dozens of years, culture - hundreds of years, and biological evolution - most likely dozens of thousands of years. The speed of increase of p that might be interpreted as a speed of cultural changes has yet another characteristic time scale and our work shows that this scale might influence the evolution of language. Certainly, further research would be needed to examine in more detail an intricate role played by learning, culture and biological evolution on language.

4 Conclusions

In the present paper we examined an evolutionary naming game model. Simulations show that coupling of linguistic and evolutionary ingredients produces a discontinuous transition and that learning can direct the evolution towards better linguistic abilities (Baldwin effect). However, under perhaps more realistic assumptions, when the communication probability increases continuously,
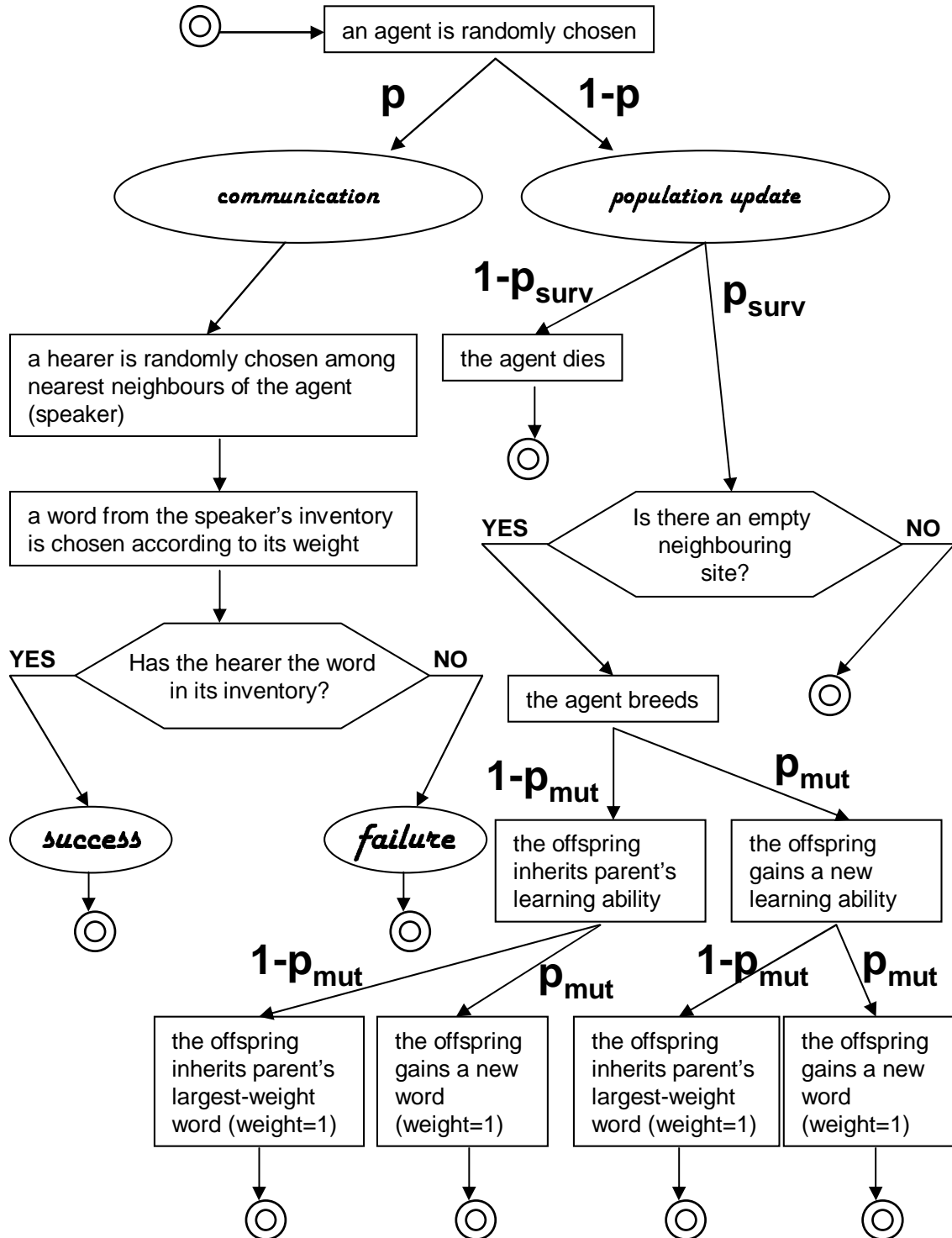
this transition is split into a series of transitions. It shows that the speed of cultural changes might be yet another factor affecting the evolution of language and setting additional characteristic timescale. The present model is not very demanding computationally . It seems to be possible to consider agents talking on more than one object, or to examine statistical properties of simulated languages such as for example, distributions of their lifetimes or of the number of users. One can also study effects like diffusion of languages, the role of geographical barriers, or formation of language families. There is already an extensive literature documenting linguistic data as well as various computational approaches modelling, for example, competition between already existing natural languages [37, 38, 39]. The dynamics of the present model, that is based on an act of elementary communication, offers perhaps more natural description of dynamics of languages than some other approaches that often use some kind of coarse-grained dynamics.

There are also more physical aspects of the proposed model that might be worth further studies. As we have already mentioned, when the learning ability is kept fixed, the transition between disordered and linguistically coherent phases seems to be continuous. On the other hand, such a transition resembles the symmetry breaking transition in the q-state Potts model, where at sufficiently low temperature the model collapses on one of the q ground states. However, in the two-dimensional case and for large q (q in our case corresponds to the number of all languages used by agents) such a transition is known to be discontinuous. Of course the dynamics of our model is much different from Glauber or Metropolis dynamics that reproduce the equilibrium Potts model, but very often such differences are irrelevant as long as, for example, the symmetry of the model is preserved (which is the case for our model). Another possibility that would explain a continuous nature of the transition in our case might be a different nature of (effective) domain walls between clusters. In our model these domain walls in some cases might be much softer and that would shift the behaviour of our model toward models with continuous-like symmetry (as e.g., XY model). To clarify this issue some further work is, however, needed.

Appendix

The following diagram illustrates an elementary step of the dynamics. The change of inventories in the case of *success* or *failure* is described in the text.

```
                    ┌─────────────────────────────┐
      ◎ ──────────► │ an agent is randomly chosen │
                    └─────────────────────────────┘
                           p          1-p

              ╭──────────────╮        ╭────────────────────╮
              │ communication │        │ population update   │
              ╰──────────────╯        ╰────────────────────╯
                                  1-p_surv        p_surv

    ┌─────────────────────────────┐   ┌───────────────┐
    │ a hearer is randomly chosen │   │ the agent dies │
    │ among nearest neighbours of │   └───────────────┘
    │ the agent (speaker)         │          ◎
    └─────────────────────────────┘

    ┌─────────────────────────────┐        ┌──────────────────┐
    │ a word from the speaker's   │   YES  │ Is there an empty │ NO
    │ inventory is chosen         │◄───────│ neighbouring site?│────►
    │ according to its weight     │        └──────────────────┘
    └─────────────────────────────┘                               ◎

   YES  ┌──────────────────────┐ NO      ┌─────────────────┐
   ◄────│ Has the hearer the   │────►    │ the agent breeds │
        │ word in its inventory?│        └─────────────────┘
        └──────────────────────┘      1-p_mut          p_mut

    ╭─────────╮      ╭─────────╮
    │ success │      │ failure │
    ╰─────────╯      ╰─────────╯
        ◎                ◎
```

The following boxes (from left to right):

- **1-p_mut** → the offspring inherits parent's learning ability
- **p_mut** → the offspring gains a new learning ability

From "the offspring inherits parent's learning ability":
- **1-p_mut** → the offspring inherits parent's largest-weight word (weight=1) → ◎
- **p_mut** → the offspring gains a new word (weight=1) → ◎

From "the offspring gains a new learning ability":
- **1-p_mut** → the offspring inherits parent's largest-weight word (weight=1) → ◎
- **p_mut** → the offspring gains a new word (weight=1) → ◎

References

1. N. Chomsky, *Aspects of the theory of syntax*, (Cambridge, MA: MIT Press, 1965).
2. G. Sampson, *Educating Eve: The 'Language Instinct' Debate*, (Cassell, 1997).
3. N. Chomsky, Language and Mind. San Diego, Harcourt Brace Jovanovich (1972).
4. S. Pinker and P. Bloom, Behav. Brain Sci. **13**, 707 (1990).
5. R. S. Jackendoff, *Languages of the mind*, (MIT Press, 1992).
6. C. Knight et al. (eds.), *The Evolutionary Emergence of Language Social Function and the Origin of Linguistic Form*, (Cambridge University Press, 2000).
7. M. A. Nowak and N. L. Komarova, Trends in Cogn. Sci. **5**, 288 (2001). M. A. Nowak and D. C. Krakauer, Proc. Natl. Acad. Sci. USA **96**, 8028 (1999).
8. K. Smith, Ph.D. thesis, University of Edinburgh (2003).
9. S. J. Gould, *The limits of adaptation: Is language a spandrel of the human brain?*, unpublished paper delivered to the Center for Cognitive Science, MIT.
10. J. Maynard Smith and E. Szathmáry, *Major Transitions in Evolution* (Freeman, 1995).
11. S. Számadó and E. Szathmáry Trends. Ecol. Evol. **21**, 555 (2006).
12. W. D. Hamilton, J. Theor. Biol. **7**, 1 (1964).
13. R. L. Trivers, Quarterly Review of Biology **46**, 35 (1971).
14. J. L. Dessalles, in J. R. Hurford et al. (eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases* (Cambridge University Press, Cambridge, 1998).
15. J. Noble, in *Evolutionary Emergence of Language*, Knight et al. (eds.) (Cambridge University Press, 2000).
16. J. M. Baldwin, Amer. Natur. **30**, 441 (1896).
17. G. G. Simpson, Evolution **7**, 110 (1953).
18. B. H. Weber and D. J. Depew (eds.), *Evolution and Learning - The Baldwin Effect Reconsidered*, (Cambridge, MA, MIT Press, 2003).
19. P. Turney, *Myths and legends of the Baldwin Effect.* In T. Fogarty and G. Venturini (Eds.), Proceedings of the ICML-96 (13th International Conference on Machine Learning, Bari, Italy).
20. H. Yamauchi, *Baldwinian Accounts of Language Evolution*, PhD thesis, The University of Edinburgh, Edinburgh, Scotland (2004).
21. M. H. Christiansen and N. Chater, *Language as shaped by the brain*, preprint (2007).
22. M. A. Nowak, N. L. Komarova, and P. Niyogi, Nature **417**, 417 (2002).
23. B. de Boer, *Computer modelling as a tool for understanding language evolution*, in *Evolutionary Epistemology, Language and Culture - A nonadaptationist system theoretical approach*, Gonthier et al. (eds.) (Dordrecht: Springer, 2006).
24. S. Kirby, Artif. Life **8**, 185 (2002).
25. L. Steels, in *Proceedings of the International Workshop of the Self-Organization and Evolution of Social Behaviour*, C. Hemelrijk and E. Bonabeau (eds.) (University of Zurich, Switzerland, 2002).
26. S. Kirby and J. Hurford, *The emergence of Linguistic Structure; An Overview of the Iterated Learning Model*, in *Simulating the Evolution of Language*, A. Cangelosi and D. Parisi (eds.) (Springer-Verlag, Berlin, 2001).
27. H. Brighton, Artif. Life **8**, 25 (2002).
28. L. Steels, Artif. Life **2**, 319 (1995).
29. A. Lipowski and D. Lipowska, Bio-linguistic transition and the Baldwin effect in the evolutionary naming game model, preprint, Int. J. Mod. Phys. C (in press).

30. A. Baronchelli, M. Felici, V. Loreto, E. Caglioti, and L. Steels, J. Stat. Mech. **P06014** (2006).
31. L. Dall'Asta, A. Baronchelli, A. Barrat, and V. Loreto, Phys. Rev. E **74**, 036105 (2006).
32. D. Nettle, Lingua **108**, 95 (1999). **ibid., 108**, 119 (1999).
33. Java applet that illustrates dynamics of our model is available at: `http://spin.amu.edu.pl/~lipowski/biolin.html` or `http://www.amu.edu.pl/~lipowski/biolin.html`
34. G. Hinton and S. Nowlan, Complex Systems **1**, 495 (1987).
35. S. Munroe and A. Cangelosi, Artif. Life **8**, 311 (2002).
36. C. Holden, Science **303**, 1316 (2004)
37. D. Abrams and S. H. Strogatz, Nature **424**, 900 (2003).
38. C. Schulze, D. Stauffer, and S. Wichmann, Commun. Comp. Phys. **3**, 271 (2008).
39. P. M. C. de Oliveira, D. Stauffer, S. Wichmann, and S. M. de Oliveira, *A computer simulation of language families*, e-print:arXiv:0709.0868.